



## Infraestruturas de investigação: o caso do Social Sciences DataLab

Ana Catarina Mateus Reis<sup>a</sup>,

<sup>a</sup>*Universidade Nova de Lisboa, Portugal, [catarina.reis@novasbe.pt](mailto:catarina.reis@novasbe.pt)*

---

### Resumo

As infraestruturas de investigação afirmam-se cada vez mais como um elemento fundamental para a construção de novo conhecimento e para a inovação. O Social Sciences DataLab, uma das infraestruturas inseridas no Roteiro Nacional das Infraestruturas de Investigação de Interesse Estratégico para o período de 2014 a 2020, pretende ser uma referência no que toca à investigação em Ciências Sociais, proporcionando acesso a um alargado conjunto de bases de dados, com um foco nas áreas de Economia, Finanças e Gestão. Resultante de uma parceria entre a Faculdade de Economia da Universidade Nova de Lisboa e a Universidade do Minho, o DataLab apoia também o projeto SHARE – Survey of Health, Ageing and Retirement in Europe. Embora o balanço do primeiro ano de funcionamento seja positivo, vários desafios se colocam ao DataLab para que este se possa afirmar como uma infraestrutura de investigação sustentável. À semelhança de outras infraestruturas nacionais e internacionais, o seu desenvolvimento futuro terá de passar pelo aumento da sua visibilidade, pela garantia de financiamento a longo prazo e pela colaboração com outros projetos e infraestruturas.

**Palavras-chave:** Infraestruturas de investigação, Social Sciences DataLab, microdados, Ciências Sociais e Humanas, Horizonte 2020

---

### Introdução e metodologia

As infraestruturas de investigação têm um importante papel na construção de novo conhecimento em todas as áreas científicas. Consideradas um pilar da inovação, infraestruturas de qualidade e acessíveis são necessárias para conduzir investigação competitiva internacionalmente, o que se tem refletido no crescente investimento a nível europeu nesta área, nomeadamente no âmbito do Programa-Quadro Horizonte 2020.

O Social Sciences DataLab é uma infraestrutura de investigação que se pretende afirmar como uma referência nas Ciências Sociais, alargando o acesso a bases de dados pela comunidade científica e assim contribuindo para melhorar as condições de investigação nesta área em Portugal.

Esta comunicação tem como principal objetivo apresentar o DataLab e situá-lo no contexto mais geral das infraestruturas de investigação. Parte-se de uma revisão da literatura relativa às infraestruturas de investigação, em particular no âmbito das Ciências Sociais e Humanas, procurando clarificar as suas principais características, bem como o seu papel e valor. De seguida apresenta-se o DataLab e faz-se o ponto de situação da infraestrutura depois de um ano de funcionamento, discutindo-se a forma como esta desempenha a sua missão, a sua relevância e os desafios que se lhe colocam no futuro.

### Resultados e discussão

O Social Sciences DataLab (Laboratório de Dados de Ciências Sociais), doravante designado por DataLab, é uma infraestrutura distribuída que resulta de uma parceria da Faculdade de Economia da Universidade Nova de Lisboa e a Universidade do Minho. É uma das quarenta infraestruturas de investigação incluídas no Roteiro Nacional das Infraestruturas de Investigação de Interesse Estratégico, para o período de 2014 a 2020, na área de Ciências Sociais e Humanidades.

Infraestruturas de investigação consistem em plataformas que fornecem recursos e serviços, utilizadas pelas comunidades científicas para desenvolver investigação de ponta em áreas específicas. Elas podem ser geograficamente concentradas (ou locais), distribuídas (como rede organizada de recursos) ou virtuais, e assumir uma grande variedade de formas, desde

“equipamento científico de grande porte, conjuntos de instrumentos científicos, coleções e outros recursos baseados no conhecimento, arquivos e dados científicos, sistemas computacionais e de programação, redes de comunicação que promovam o acesso aberto digital, bem como outras infraestruturas de natureza única essenciais para atingir a excelência na investigação e na inovação” (Fundação para a Ciência e a Tecnologia, 2014, p.9).

Apesar desta grande diversidade, é possível identificar um conjunto de características comuns a todas as infraestruturas de investigação: elas oferecem recursos e serviços como um bem público, procuram dar resposta às necessidades da comunidade científica, são duráveis e estáveis a longo-prazo, adaptam-se continuamente às necessidades dos seus utilizadores, e contribuem para uma maior qualidade e eficiência da investigação que apoiam (Renschler, Kleiner, Wernli, 2013, p.15-16).

De grande importância em todas as áreas científicas, as infraestruturas de investigação estão a ganhar cada vez mais relevo nas Ciências Sociais e Humanas, o que tem efeitos profundos na forma como a investigação nestas áreas é organizada e conduzida (Kleiner et al., 2013, p.194).

Um dos impactos mais relevantes das infraestruturas de investigação é o facto de estas contribuírem para desfazer barreiras no acesso à informação. Através delas, os investigadores podem aceder facilmente a grandes volumes dados, numa multiplicidade de tópicos, sendo que, em muitos casos, este acesso pode ser feito independentemente da localização física das próprias infraestruturas. Cada vez mais os dados são disponibilizados *online*, de forma gratuita, requerendo apenas a autenticação numa plataforma virtual. Um outro aspeto crucial é a forma como as infraestruturas de investigação facilitam a criação de parcerias e a cooperação, muitas vezes a nível internacional, em projetos comuns. Elas disponibilizam plataformas técnicas que agilizam a troca de dados e fomentam a criação de redes de colaboração e partilha de recursos e conhecimentos, contribuindo para a disseminação de boas práticas.

Deste modo, é possível afirmar que as infraestruturas de investigação são uma poderosa ferramenta na construção de uma ciência mais aberta. Elas têm promovido, ao longo das últimas décadas, um maior acesso à informação e o desenvolvimento de um modelo mais colaborativo de fazer ciência.

As infraestruturas de investigação têm um lugar importante na estratégia europeia, nomeadamente no âmbito do Horizonte 2020 – Programa-Quadro Comunitário de Investigação & Inovação. Este programa financeiro da União Europeia (que irá ter como sucessor o Horizonte Europa) decorre entre 2014 e 2020, e é o maior instrumento da Comunidade Europeia especificamente orientado para o apoio à investigação, através do cofinanciamento de projetos de investigação. O apoio financeiro concedido reflete as prioridades da estratégia Europa 2020 e, em particular, da iniciativa Innovation Union, que tem como objetivo assegurar a competitividade global da Europa e assenta em três pilares: excelência científica, liderança industrial e desafios sociais. Para a excelência científica, é afirmada a

necessidade criar e manter infraestruturas de investigação de elevada qualidade (Comissão Europeia, 2014, p.5, 7-8).

A investigação em Ciências Sociais e Humanas tem também um lugar de destaque no Horizonte 2020, tendo sido integrada em todos os objetivos específicos do programa. Isto reflete o reconhecimento do contributo essencial das Ciências Sociais e Humanas, no âmbito de uma abordagem multidisciplinar, para dar resposta aos desafios que a Europa contemporânea enfrenta, tais como as mudanças demográficas, a segurança alimentar, as migrações, as alterações climáticas ou a inclusão social (Duşa, Oellers, Wolff, 2014, p.225).

O projeto SHARE – Survey of Health, Ageing and Retirement in Europe é um exemplo do investimento neste campo. Criada em 2004, trata-se uma infraestrutura de investigação distribuída, pan-europeia, coordenada pelo Munich Center for the Economics of Aging (MEA), no Max Planck Institute for Social Law and Social Policy. Este projecto multi-disciplinar faz a recolha e tratamento de dados comparáveis sobre o envelhecimento na Europa, nomeadamente relativos a saúde, situação socioeconómica e redes sociais e familiares de indivíduos com idade igual ou superior a 50 anos, de 20 países europeus e Israel. Esta infraestrutura tem o seu pólo nacional na Universidade do Minho e é apoiada pelo DataLab. Os dados recolhidos são disponibilizados *online* e facultados gratuitamente a investigadores que estejam integrados em instituições científicas, os quais deverão assinar uma declaração em que se comprometem a utilizar os dados exclusivamente para fins científicos.

Em 2006, o SHARE foi selecionado como um dos projetos a ser implementados no Fórum Estratégico Europeu para as Infraestruturas de Investigação (European Strategy Forum on Research Infrastructures - ESFRI). Criado pela Comissão Europeia em 2002, o ESFRI é responsável por estabelecer uma estratégia europeia para as infraestruturas de investigação que estimule a competitividade internacional da investigação realizada na Europa. Para levar a cabo a sua estratégia, o ESFRI apoia uma rede de infraestruturas de investigação de interesse europeu a longo prazo, suportando projetos em diversas áreas. Em 2011, o SHARE adquiriu um novo estatuto legal, constituindo-se como o primeiro European Research Infrastructure Consortium – ERIC, quadro legal que visa facilitar o estabelecimento e operação de infraestruturas de interesse europeu com o envolvimento de diversos países europeus.

Instrumentos como o ESFRI ou o estatuto de ERIC ajudam a fazer face a um dos desafios mais prementes que se colocam às infraestruturas de investigação: a questão da sua sustentabilidade. Na maior parte dos casos, as infraestruturas de investigação necessitam de um grande investimento inicial e/ou elevados custos de manutenção. Assim, o seu funcionamento a longo-prazo está dependente da existência de financiamento adequado, ou seja, suficiente para fazer face às suas necessidades e, simultaneamente, estável e de longa-duração. A estabilidade no funcionamento das infraestruturas é fundamental para que o investimento feito possa ter o devido retorno. A coordenação a nível europeu na área das infraestruturas é importante a este nível, pois permite partilhar os custos envolvidos na instalação e operação, que em certas infraestruturas são tão elevados que não estão ao alcance de um único país ou organismo financiador. A coordenação entre diferentes infraestruturas deve ser feita tendo em vista também um uso mais eficiente dos recursos, na medida em que evita a duplicação de infraestruturas com propósitos semelhantes.

Um outro fator determinante para a continuidade das infraestruturas de investigação consiste na sua capacidade para se manterem atualizadas e relevantes. Para isso elas devem procurar ativamente acompanhar os desenvolvimentos das áreas científicas onde atuam e alinhar-se com as necessidades dos seus utilizadores.

As infraestruturas de investigação alojam frequentemente dados sensíveis, que necessitam de cuidados especiais quer a nível do seu armazenamento, quer da consulta e partilha. Assim, outro dos desafios que se coloca às infraestruturas de investigação consiste em conciliar o acesso aberto e partilha de informação com as exigências ao nível da privacidade e proteção de dados.

Por último, é importante que as infraestruturas de investigação se empenhem em aumentar a sua visibilidade junto de investigadores, financiadores e público em geral. É necessário encorajar os investigadores a incorporarem as infraestruturas no seu trabalho, divulgando os recursos e serviços oferecidos, e tentando ao máximo eliminar dificuldades, técnicas e outras, à utilização das infraestruturas. Importa medir o impacto e comunicar a relevância da investigação realizada com base nelas, salientando os seus benefícios para a sociedade (Duşa, Oellers, Wolff, 2014, p.231-232).

A inclusão do DataLab no Roteiro Nacional das Infraestruturas de Investigação de Interesse Estratégico resultou da candidatura submetida ao concurso lançado em 2013 pela Fundação para a Ciência e a Tecnologia (FCT) para a criação de um roteiro nacional alinhado com o ESFRI. Para serem incluídas nos roteiros nacionais e europeus, as infraestruturas de investigação devem cumprir um conjunto de requisitos. Elas devem ter uma “gestão profissionalizada, que garanta a implementação de um plano de ação e o cumprimento dos objetivos aí definidos, fazendo uma gestão eficiente e transparente dos recursos”; devem ser capazes “de se relacionar e prestar serviços às comunidades científica, educacional, empresarial e industrial”; e devem, por último, “ter as condições de acesso por parte de investigadores nacionais e estrangeiros externos à infraestrutura bem definidas e devidamente publicitadas, e integradas nos objetivos e no plano de ação da infraestrutura de investigação” (Fundação para a Ciência e a Tecnologia, 2014, p.9). O financiamento atribuído pela FCT destina-se à construção ou adaptação de infraestruturas físicas, aquisição de instrumentos e equipamento científico e técnico e despesas com recursos humanos.

O DataLab teve início em Junho de 2017 e a sua estratégia de implementação passou por um aproveitamento dos recursos e competências já existentes na Faculdade de Economia, nomeadamente na Unidade de Investigação, Biblioteca e instalações de Tecnologias da Informação, e uma identificação das necessidades para a consecução do projeto. Com o financiamento obtido foi possível contratar uma equipa dedicada ao projeto e subscrever vários recursos eletrónicos que se juntaram aos que a Biblioteca já possuía.

O DataLab disponibiliza aos seus utilizadores, de forma gratuita, um conjunto de recursos eletrónicos essenciais para conduzir investigação avançada em Ciências Sociais, com um foco nas áreas de Economia, Finanças e Gestão. Os recursos oferecidos consistem em bases de dados especializadas que dão acesso a literatura e dados sociais, económicos e financeiros. Um dos princípios do DataLab é alargar o acesso aos recursos para lá da academia. Assim, procurou-se salvaguardar, aquando da sua subscrição destas bases de dados, a possibilidade de utilizadores externos à Faculdade poderem aceder às mesmas, com autorização prévia.

Através da realização de protocolos com instituições públicas, o DataLab oferece também à comunidade de investigadores acesso a conjuntos de microdados. Por microdados entende-se dados que correspondem ao nível mais elementar de desagregação da informação estatística, ou seja, dados estatísticos individuais sobre pessoas singulares e coletivas. Estes dados são muitas vezes recolhidos no âmbito de processos administrativos, no entanto, podem servir outros propósitos não previstos na altura da sua recolha. Dar acesso a estes dados para investigação “é uma forma económica e eficiente de multiplicar e diversificar a análise e uso de informação já existente”, oferecendo “oportunidades inesgotáveis de geração de novo conhecimento” (Dupriez, Boyko, 2010, p.1).

Dados administrativos são particularmente interessantes para a realização de trabalhos de investigação em Ciências Sociais, quer pela sua cobertura e detalhe, quer pela sua atualidade, além de muitas vezes estarem já prontos para serem analisados (Duşa, Oellers, Wolff, 2014, p.230). Os microdados dão oportunidades de investigação de ponta em diversas áreas, sendo que a sua importância vai para além da academia, já que eles podem ser usados, por exemplo, para definir e avaliar políticas públicas tendo por base evidência científica. Por serem dados particularmente sensíveis, os microdados exigem cuidados especiais no que toca ao seu armazenamento e acesso, razão pela qual o DataLab dispõe de uma sala – Safe Centre – utilizada exclusivamente para esse efeito. O acesso aos microdados requer autorização prévia e é feito em equipamento informático dedicado, sem qualquer ligação a meios de gravação ou transferência de informação. O acesso ao Safe Centre é restrito e controlado pela equipa do DataLab.

De momento, o DataLab dá acesso, através de um protocolo com o Gabinete de Estratégia e Planeamento do Ministério do Trabalho, Solidariedade e Segurança Social, aos “Quadros de Pessoal”, que disponibilizam informação sobre estrutura empresarial, emprego, duração do trabalho e remunerações em Portugal, desde 1981. Existe também um protocolo com a Direção-Geral de Estatísticas da Educação e Ciência do Ministério da Educação e Ciência para dar acesso à base de dados “MISI Anonimizada”, que fornece informação sobre a educação pré-escolar e dos ensinos básico e secundário. A informação é cedida exclusivamente para fins de investigação, sem intuito comercial, e salvaguardando o segredo estatístico e a proteção de dados. No futuro, pretendemos celebrar mais protocolos para facilitar o acesso dos investigadores a dados administrativos, contribuindo para promover a sua utilização para a realização de estudos.

As infraestruturas de investigação não se limitam a oferecer acesso a dados. Elas disponibilizam também ferramentas, aconselhamento aos seus utilizadores e formação. Esta última, em particular, é importante quer para sensibilizar os utilizadores para os recursos que a infraestrutura oferece, quer para promover boas práticas na utilização dos dados e da informação, técnicas e procedimentos inovadores. Nesta linha, além de disponibilizar o acesso a recursos de investigação, o DataLab organiza também sessões de formação abertas a toda a comunidade relacionadas com esses mesmos recursos, tais como a realização de pesquisa ou o tratamento de dados. É também dada assistência aos investigadores em questões relacionadas com o acesso, pesquisa e publicação de trabalhos baseados nos recursos disponibilizados pelo DataLab.

Informação sobre o projeto é publicada no seu *website* oficial. Nele os utilizadores podem encontrar uma lista dos recursos eletrónicos disponibilizados, consultar o calendário de formação e aceder ao formulário de pedido de acesso a bases de dados e microdados (para utilizadores externos), e formulário de contato. Podem ainda comunicar com um membro da equipa do DataLab para colocar questões através de um *chat*. São também usadas as redes sociais para divulgar formações e disponibilizar informação sobre os recursos disponíveis. A fim de obter *feedback*, são feitos inquéritos de satisfação aos participantes nas sessões de formação. Os professores e investigadores da Faculdade de Economia são inquiridos periodicamente acerca das suas necessidades em termos de recursos de investigação, para que os recursos disponibilizados correspondam às suas necessidades.

Embora o balanço do primeiro ano de funcionamento seja positivo, vários desafios se colocam ao DataLab para que este se possa afirmar como uma infraestrutura de investigação de referência e com continuidade.

Em primeiro lugar, será necessário garantir o financiamento da infraestrutura a longo prazo, de modo a que possamos continuar a oferecer recursos e serviços relevantes e de qualidade, e ter uma equipa afeta ao projeto. Os projetos de investigação duram, por vezes, vários anos, pelo que é necessário assegurar que os investigadores continuam a ter acesso às bases de dados de que necessitam. O

financiamento a longo prazo é, assim, fundamental para que o investimento feito na infraestrutura possa ter o devido retorno, traduzido num aumento da qualidade e quantidade dos trabalhos de investigação produzidos.

Em segundo lugar, terá de ser desenvolvido um esforço para aumentar a visibilidade do DataLab, o que passará por um melhor trabalho ao nível da comunicação, nomeadamente através da melhoria do *website* do projeto, que deverá incluir mais informação sobre os recursos e as formas de acesso e condições de utilização dos mesmos. O DataLab pretende alcançar um maior número de pessoas, sendo que uma maior aposta nas redes sociais pode ajudar bastante neste aspeto. Pretende-se também criar um plano de formação, em conjunto com a Biblioteca, para promover uma maior utilização dos recursos do DataLab. Desenvolver uma melhor comunicação com os nossos utilizadores é muito importante para perceber quais as suas necessidades de informação e formação, permitindo ajustar o orçamento disponível para subscrever recursos e oferecer serviços mais relevantes. Uma monitorização das estatísticas de utilização dos recursos, pedidos de acesso e trabalhos de investigação publicados (realizados com base nos recursos oferecidos) são exemplos de indicadores que nos ajudam a compreender o impacto e relevância do projeto.

Por último, a colaboração do DataLab com outros projetos e infraestruturas, incluindo criação de novas parcerias para dar acesso a mais recursos, em particular, microdados, será outro passo importante para que ele possa cumprir a sua missão.

### **Conclusões**

O DataLab pretende ser uma infraestrutura de referência na área das Ciências Sociais, fornecendo recursos e serviços que vão ao encontro das necessidades dos investigadores e contribuindo para uma maior qualidade e eficiência da investigação, para a construção de novo conhecimento e inovação. O uso de bases de dados especializadas e de microdados proporciona um conhecimento mais completo e profundo dos aspetos sociais, culturais, políticos e económicos da sociedade contemporânea. Quando utilizados para definir e avaliar políticas públicas, eles contribuem para tomadas de decisão mais informadas. Em linha com outras infraestruturas de investigação, o DataLab oferece não só recursos, mas também apoio e formação. A recolha regular de *feedback* acerca dos recursos e serviços permite adaptar a oferta às necessidades dos utilizadores e, juntamente com outros indicadores, medir o impacto da infraestrutura. Tal como outras infraestruturas nacionais e internacionais, o DataLab depara-se com alguns desafios, entre os quais garantir o financiamento da infraestrutura a longo prazo, aumentar a sua visibilidade, colaborar com outros projetos e infraestruturas, e desenvolver novas parcerias. Apesar dos elevados custos de investimento envolvidos, espera-se que o desenvolvimento da infraestrutura, ao permitir o acesso a um grande conjunto de dados e à sua reutilização em diferentes projetos, resulte numa redução de custos e um elevado retorno para a investigação.

### **Referências**

COMISSÃO EUROPEIA (2014) – *Horizon 2020 em breves palavras*. [Em linha]. Luxemburgo: Serviço das Publicações Oficiais das Comunidades Europeias. [Consult. 24 Jul. 2018]. Disponível na Internet:

<URL:[https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/H2020\\_PT\\_KI0213413PTN.pdf](https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/H2020_PT_KI0213413PTN.pdf)>. ISBN 978-92-79-38928-3.

COMISSÃO EUROPEIA. Horizon Europe - the next research and innovation framework programme. [Em linha]. [Consult. 24 Jul. 2018]. Disponível na Internet: <URL: [https://ec.europa.eu/info/designing-next-research-and-innovation-framework-programme/what-shapes-next-framework-programme\\_en](https://ec.europa.eu/info/designing-next-research-and-innovation-framework-programme/what-shapes-next-framework-programme_en)>

COMISSÃO EUROPEIA. What is Horizon 2020?. [Em linha]. [Consult. 24 Jul. 2018]. Disponível na Internet: <URL: <https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>>.

DUPRIEZ, Olivier; BOYKO, Ernie (2010) – Dissemination of Microdata Files. Formulating Policies and Procedures. *International Household Survey Network* [Em linha]. IHSN Working Paper No 005. [Consult. 24 Jul. 2018]. Disponível na Internet: <URL: <http://ihsn.org/sites/default/files/resources/IHSN-WP005.pdf> >.

DUŞA, Adrian; OELLERS, Claudia; WOLFF, Simon (2014) – A Common Agenda for the European Research Infrastructures in the Social Sciences and Humanities. In *Facing the Future: European Research Infrastructures for the Humanities and Social Sciences*. Berlin : SCIVERO. ISBN 978-3-944417-03-5.

FUNDAÇÃO PARA A CIÊNCIA E A TECNOLOGIA (2014) – *Portuguese Roadmap for Research Infrastructures 2014–2020*. [Em linha]. Lisboa : Fundação para a Ciência e a Tecnologia. [Consult. 23 Jul. 2018]. Disponível na Internet: <URL: <https://www.fct.pt/apoios/equipamento/roteiro/index.phtml.pt>>.

KLEINER, Brian [et al.] (2013) – Impact, challenges, and outlook: How research infrastructures are changing the social sciences. In *Understanding Research Infrastructures in the Social Sciences*. Zurich : Seismo Press. ISBN 978-3-03777-133-4.

RENSCHLER, Isabelle; KLEINER, Brian; WERNLI, Boris (2013) – Concepts and key features for understanding social science research infrastructures. In *Understanding Research Infrastructures in the Social Sciences*. Zurich : Seismo Press. ISBN 978-3-03777-133-4.

SHARE. Início. *SHARE - Survey of Health, Ageing and Retirement in Europe*. [Em linha]. [Consult. 23 Jul. 2018]. Disponível na Internet: <URL: <http://www.share-project.pt/>>.